

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
"Амурский государственный университет"

УТВЕРЖДАЮ

Проректор по учебной и научной
работе

Лейфа А.В. Лейфа

19 июня 2024 г.

РАБОЧАЯ ПРОГРАММА
«СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЧИ»

Направление подготовки 45.04.03 Фундаментальная и прикладная лингвистика

Направленность (профиль) образовательной программы – Иностранные языки и речевые
технологии

Квалификация выпускника – Магистр

Год набора – 2024

Форма обучения – Очная

Курс 2 Семестр 3

Зачет 3 сем

Общая трудоемкость дисциплины 72.0 (академ. час), 2.00 (з.е)

Составитель С.В. Андросова, профессор, д-р филол. наук

Филологический факультет

Кафедра иностранных языков

Рабочая программа составлена на основании Федерального государственного образовательного стандарта ВО для направления подготовки 45.04.03 Фундаментальная и прикладная лингвистика, утвержденного приказом Министерства науки и высшего образования Российской Федерации от 12.08.20 № 993

Рабочая программа обсуждена на заседании кафедры иностранных языков

01.02.2024 г. , протокол № 6

Заведующий кафедрой Морозова О.Н. Морозова

СОГЛАСОВАНО

Учебно-методическое управление

Чалкина Н.А. Чалкина

19 июня 2024 г.

СОГЛАСОВАНО

Выпускающая кафедра

Морозова О.Н. Морозова

19 июня 2024 г.

СОГЛАСОВАНО

Научная библиотека

Петрович О.В. Петрович

19 июня 2024 г.

СОГЛАСОВАНО

Центр цифровой трансформации и
технического обеспечения

Тодосейчук А.А. Тодосейчук

19 июня 2024 г.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель дисциплины:

– дать многоаспектное описание методов статистического анализа и возможностей их применения для лингвистических исследований.

Задачи дисциплины:

- ознакомить с основными статистическими методами и возможностями их применения в различных областях лингвистике;
- детально рассмотреть алгоритмы процедур статистической обработки полученных в ходе исследования результатов;
- выработать навыки правильного отбора и применения соответствующих алгоритмов для собственных лингвистических исследований.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Статистическая обработка речи» относится к дисциплинам формируемыми участниками образовательных отношений. В качестве пререквизитов для успешного ее освоения предполагает формирование знаний, умений и навыков, связанных с накоплением и организацией лингвистических данных, а также с информационными технологиями.

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ИНДИКАТОРЫ ИХ ДОСТИЖЕНИЯ

3.1 Профессиональные компетенции и индикаторы их достижения

Код и наименование профессиональной компетенции	Код и наименование индикатора достижения профессиональной компетенции
ПК-2. Способен разрабатывать лингвистические технические документы, адресованные специалисту по информационным технологиям, описывать информационные и математические модели, технические решения с точки зрения специалиста по информационным технологиям	ИД-1 Знает основные принципы и методы разработки лингвистической технической документации, способы описания информационных и математических моделей и технических решений. ИД-2 Умеет разрабатывать лингвистические технические документы, адресованные специалисту по информационным технологиям, описывать информационные и математические модели, технические решения с точки зрения специалиста по информационным технологиям. ИД-3 Владеет практическим опытом создания технических документов, описания математических моделей и технических решений в рамках своей профессиональной деятельности.
ПК-3. Способен планировать и организовывать аналитические работы в языковом информационно-технологическом проекте, управлять ими	ИД-1 Знает основные принципы и этапы планирования и организации аналитических работ в проектах информационно-технологической направленности, способы управления аналитическими работами. ИД-2 Умеет планировать и организовывать аналитические работы в языковом информационно-технологическом проекте, управлять ими.

	ИД-3 Владеет практическими навыками планирования и организации аналитических работ в проектах информационно- технологического этапа в рамках своей профессиональной деятельности.
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

4. СТРУКТУРА ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 2.00 зачетных единицы, 72.0 академических часов.

1 – № п/п

2 – Тема (раздел) дисциплины, курсовая работа (проект), промежуточная аттестация

3 – Семестр

4 – Виды контактной работы и трудоемкость (в академических часах)

4.1 – Л (Лекции)

4.2 – Лекции в виде практической подготовки

4.3 – ПЗ (Практические занятия)

4.4 – Практические занятия в виде практической подготовки

4.5 – ЛР (Лабораторные работы)

4.6 – Лабораторные работы в виде практической подготовки

4.7 – ИКР (Иная контактная работа)

4.8 – КТО (Контроль теоретического обучения)

4.9 – КЭ (Контроль на экзамене)

5 – Контроль (в академических часах)

6 – Самостоятельная работа (в академических часах)

7 – Формы текущего контроля успеваемости

1	2	3	4									5	6	7
			4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9			
1	Статистический метод	3	1		7								9.45	Устный опрос, тест, практическое задание. Тест с практическим заданием, портфолио.
2	Элементарные инструменты наблюдения за действием статистических законов	3	1		7								9.45	Устный опрос, тест, практическое задание с использованием информационных технологий. Тест с практическим заданием, портфолио
3	Нормирование	3	3		7								9.45	Устный

	и дальнейшая статистическая обработка показателей												опрос, тест, практическое задание с использованием информационных технологий, круглый стол. Тест с практическим заданием, портфолио
4	Сравнение долей, средних выборочных частот и частотных рядов	3	1		7							9.45	Устный опрос, тест, практическое задание с использованием информационных технологий, круглый стол. Тест с практическим заданием, портфолио.
5	Зачет	3							0.2				
	Итого		6.0		28.0		0.0	0.0	0.2	0.0	0.0	37.8	

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

5.1. Лекции

№ п/п	Наименование темы (раздела)	Содержание темы (раздела)
1	Статистический метод	<p>Субъект и объект исследования. Объект и метод исследования. Объект и знания об объекте. Позиция исследователя: объективность и беспристрастность, прагматичность. Процедурность статистического метода. Приближенность статистического метода как следствие познавательной установки на прагматичность. Системность, стохастичность, симптоматичность, многомерность, конкретность, массовость и количественность объекта исследования.</p> <p>Характеристики статистического знания. Строгость, количественность и приближенность. Прагматичность. Системность, стохастичность, симптоматичность. Многомерность, конкретность, массовость. Верифицируемость и объективность. Классификация статистических методов.</p> <p>Описательные методы и виды используемых статистик. Метод моделирования и виды используемых распределений. Методы</p>

		<p>оценивания: методы оценивания неизвестных параметров распределения и методы проверки статистических гипотез. Генеральная совокупности и методы ее оценивания на больших и малых выборках. Большие и малые совокупности, представленные собирательными множествами, понятие представительной выборки. Непараметрические и устойчивые методы: дисперсионный и другие виды анализа.</p>
2	<p>Элементарные инструменты наблюдения за действием статистических законов</p>	<ol style="list-style-type: none"> 1. Статистические распределения (ряды) и их теоретическая значимость. 2. Вероятностная теория – лингвистическая реальность. Эмпирические ряды и теоретические распределения. Принципы выделения конкретных семейств распределений. Моделирующие и критериальные распределения. 3. Гауссовость – негауссовость. 4. Типичность – нетипичность. 5. Ранг – частота. Способы задания функциональной связи между случайной величиной и вероятностью. Виды рядов и соотосимые с ними лингвостатистические термины. 6. Разнообразие – ограничение разнообразия. Типы количественной вариативности лингвистических единиц. Понятие знакотипа. Построение распределения знакотипов. 7. Строение – поведение. Категории варьирующих признаков. Статусные и структурные ряды. 8. Элемент – совокупность элементов. Однопредметное и многопредметное наблюдение и виды распределения, связанные с ними: горизонтальные и вертикальные распределения. 9. Виртуальность – актуальность. Распределения в сфере фиксации и функционирования. 10. Однородность – неоднородность. Количественная и качественная однородность. Условие однородности в статистике. Нормальное распределение как основной критерий количественной однородности. Распределения с априорно устранимой и неустранимой неоднородностью. Однородность- неоднородность с точки зрения близости распределений в разных группах. Текст как статистическая совокупность. Предположение о статистической однородности текста. Показатели однородности. Критерии сравнения статистических данных. Критерий сравнения χ^2 при равной и неравной длине текста. Критерий сравнения долей языковых явлений. 11. Устойчивость – неустойчивость. 12. Редкость – частость. 13. Симметрия – асимметрия. 14. Одновершинность – многовершинность.

		15. Статика – динамика. Понятие средней величины. Использование средних величин в лингвистических исследованиях. Частота. Виды частот. Их содержательный смысл. Понятие вариационного ряда. Основные характеристики вариационного ряда. Определение длины текста. Использование статистических таблиц.
3	Нормирование и дальнейшая статистическая обработка показателей	Понятие генеральной совокупности выборки. Репрезентативность выборки. О точности и надёжности результатов наблюдения. Абсолютная и относительная ошибка выборки. Решение основных типов задач при выборочном наблюдении: определение точности результатов наблюдения, определение надёжности результатов наблюдения, определение объёма выборки при заданных t и \square . Относительная ошибка частоты фонетического явления; использование её в лингвистических исследованиях.
4	Сравнение долей, средних выборочных частот и частотных рядов	Средние и предельные величины. Стандартное квадратическое отклонение. Линейные отношения. Вычисление коэффициента корреляции и виды корреляционной связи. Метод корреляции Пирсона. Нелинейные отношения. Коэффициент ковариации. Факторный анализ.

5.2. Практические занятия

Наименование темы	Содержание темы
Статистический метод	<ol style="list-style-type: none"> 1. Определение статистического метода. Понятие о статистических данных. Соотношение математического и нематематического в статистическом методе. 2. Познавательные черты статистического метода. 3. Характеристики объекта исследования. 4. Характеристики статистического знания. 5. Основные методы сбора, систематизации, обработки и интерпретации статистических данных. 6. Описательные методы и виды используемых статистик. 7. Метод моделирования и виды используемых распределений. 7. Методы оценивания: методы оценивания неизвестных параметров распределения и методы проверки статистических гипотез. 9. Методы оценивания генеральной совокупности на больших и малых выборках. 10. Непараметрические и устойчивые методы: дисперсионный и другие виды анализа.
Элементарные инструменты наблюдения за действием статистических законов	<ol style="list-style-type: none"> 1. Статистические распределения (ряды) и их теоретическая значимость. 2. Вероятностная теория – лингвистическая реальность. 3. Гауссовость – негауссовость.

	<p>4. Типичность – нетипичность. 5. Ранг – частота. 6. Разнообразие – ограничение разнообразия. 7. Строение – поведение. 8. Элемент – совокупность элементов. 9. Виртуальность – актуальность. 10. Однородность – неоднородность. 11. Устойчивость – неустойчивость. 12. Редкость – частость. 13. Симметрия – асимметрия. 14. Одновершинность – многовершинность. 15. Статика – динамика.</p>
<p>Нормирование и дальнейшая статистическая обработка показателей</p>	<p>1. Понятие генеральной совокупности выборки. Репрезентативность выборки. 2. Точности и надёжности результатов наблюдения. Абсолютная и относительная ошибка выборки. 3. Решение основных типов задач при выборочном наблюдении: определение точности результатов наблюдения, определение надёжности результатов наблюдения, определение объёма выборки при заданных t и Δ. 4. Относительная ошибка частоты языкового явления; использование её в лингвистических исследованиях. 5. Частотность фонем, аллофонов в устной речи и модификаций. 6. Статистическая обработка длительностей звуков: среднее, среднее максимально и минимальное, стандартное квадратичное отклонение, корреляция и ковариация в структурах ГСГ. 7. Грамматический, синтаксический, семантический, стилистический анализ текстов с использованием инструментария лингвостатистики.</p>
<p>Сравнение долей, средних выборочных частот и частотных рядов</p>	<p>1. Средние и предельные величины. Стандартное квадратическое отклонение. 2. Линейные отношения. Вычисление коэффициента корреляции и виды корреляционной связи. Метод корреляции Пирсона. 3. Нелинейные отношения. Коэффициент ковариации. 4. Факторный анализ. 5. Практическое задание: используя предложенные данные вычислите коэффициенты корреляции и ковариации длительности согласного с предшествующим и последующим гласным. Проанализируйте характер и степень связи между объектами.</p>

6. САМОСТОЯТЕЛЬНАЯ РАБОТА

№ п/п	Наименование темы (раздела)	Содержание темы (раздела)	Трудоёмкость в академических
-------	-----------------------------	---------------------------	------------------------------

			часах
1	Статистический метод	Выполнение практических заданий с использованием информационных технологий, подготовка к тесту, устному опросу, создание портфолио.	9.45
2	Элементарные инструменты наблюдения за действием статистических законов	Выполнение практических заданий с использованием информационных технологий, подготовка к тесту, устному опросу, создание портфолио.	9.45
3	Нормирование и дальнейшая статистическая обработка показателей	Выполнение практических заданий с использованием информационных технологий, подготовка к тесту, устному опросу, создание портфолио, подготовка к круглому столу.	9.45
4	Сравнение долей, средних выборочных частот и частотных рядов	Выполнение практических заданий с использованием информационных технологий, подготовка к тесту, устному опросу, создание портфолио, подготовка к круглому столу.	9.45

7. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Ситуационный анализ (case-study) -- Темы 1-3

8. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания, типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков, характеризующих этапы формирования компетенций, а так же методические материалы, определяющие процедуры оценивания знаний, умений и навыков отражены в фонде оценочных средств по дисциплине.

В процессе изучения дисциплины осуществляется текущий и промежуточный контроль знаний.

Промежуточный контроль проводится в виде зачета за весь курс обучения.

Вопросы к зачету:

1. Статистика как наука: история возникновения, определение, разделы статистики, проверка достоверности утверждения, константы и переменные, количественные и качественные переменные и границы между ними. Охарактеризуйте возможные статистические параметры вашего объекта исследования.
2. Основания применения статистики для вашего лингвистического объекта, связанные с количественными характеристиками языка и речи.
3. Минимально необходимые статистические понятия: тип данных; статистический закон (динамические и статистические); вероятность и их свойства. Как их можно применить для вашего объекта исследования.
4. Нормирование отклонений (вычисление Z- показателя). Применение для вашего объекта исследования.
5. Определение значимости среднего балла (вычисление T- показателя). Применение для вашего объекта исследования.
6. Вероятностная ошибка в определении средней частоты (с использованием коэффициентов Стьюдента). Применение для вашего объекта исследования.
7. Вычисление ковариации и корреляции. Коэффициент корреляции Пирсона. Коэффициент корреляции Крамера. Интерпретация зависимости исследуемых параметров. Применение для вашего объекта исследования.

8. Распределение плотности вероятности и проверка на гауссовость / негауссовость (= хи-квадрат критерий = критерий согласия Пирсона). Применение для вашего объекта исследования.
9. Вычисление доли изучаемого явления в общей совокупности явлений. Применение для вашего объекта исследования.
10. Сравнение долей для двух рядов данных. Применение для вашего объекта исследования.
11. Вычисление стандартного квадратичного отклонения для двух рядов данных. Применение для вашего объекта исследования.
12. Вычисления с использованием критерия Стьюдента для независимых данных. Применение для вашего объекта исследования.
13. Вычисления с использованием критерия Стьюдента для зависимых выборок. Применение для вашего объекта исследования.

9. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

а) литература

1. Андросова С. В., Андросов Е. Ю. Статистическая обработка данных звучащей речи : учеб. пособие [Электронный ресурс]. -- Благовещенск : Изд-во Амур. гос. ун-та, 2016. - 60 с. (http://irbis.amursu.ru/DigitalLibrary/AmurSU_Edition/7418.pdf) (дата обращения: 17.06.2024)
2. Дегтярева, И. Н. Теория статистики : учебник / И. Н. Дегтярева. — Москва : Ай Пи Ар Медиа, 2021. — 248 с. — ISBN 978-5-4497-1212-7. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/109498.html> (дата обращения: 17.06.2024). — Режим доступа: для авторизир. пользователей
3. Полякова В.В. Основы теории статистики [Электронный ресурс] : учебное пособие / В.В. Полякова, Н.В. Шаброва. — Электрон. текстовые данные. — Екатеринбург: Уральский федеральный университет, 2015. — 148 с. — 978-5-7996-1520-8. — Режим доступа: <http://www.iprbookshop.ru/68366.html>, по паролю (дата обращения: 17.06.2024)

б) программное обеспечение и Интернет-ресурсы

№	Наименование	Описание
1	Fedora Workstation 27	Бесплатное распространение по лицензии GNU GPL http://www.gnu.org/licenses/old-licenses/gpl-2.0.htm .
2	LibreOffice	Бесплатное распространение по лицензии GNU LGPL https://ru.libreoffice.org/about-us/license/
3	Автоматизированная информационная библиотечная система «ИРБИС 64»	Лицензия коммерческая по договору №945 от 28 ноября 2011 года.
4	https://haskinslabs.org/	Сайт на английском языке свободного доступа со статьями о применении статистики в фонетических исследованиях
5	https://scholar.google.ru/	Google Scholar — поисковая система по полным текстам научных публикаций всех форматов и дисциплин.
6	http://www.philology.ru/	Philology.ru. Филологический портал. Содержит систематизированную информацию, по теоретической и прикладной науке. Центральным разделом портала является библиотека филологических текстов (монографий, статей, методических пособий).

7	http://www.mitpressjournals.org/toc/coli/37/4	Computational Linguistics — online журнал по компьютерной лингвистике
---	-----------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------

в) профессиональные базы данных и информационные справочные системы

№	Наименование	Описание
1	https://elibrary.ru/	Научная электронная библиотека eLIBRARY.RU – российский информационно-аналитический портал в области науки, технологии, медицины и образования
2	http://www.ruscorpora.ru	Национальный корпус русского языка. Информационно-справочная система, основанная на собрании русских текстов в электронной форме
3	http://neicon.ru	Полнотекстовый архив ведущих западных научных журналов на российской платформе Национального электронно-информационного консорциума (НЭИКОН)
4	http://webofscience.com	Политематическая реферативно-библиографическая и наукометрическая (библиометрическая) база данных «Web of Science Core Collection»
5	https://www.scopus.com	Международная реферативная база данных научных изданий Scopus
6	http://www.gumer.info/bibliotek_Buks/Linguist/Index_Ling.php	Библиотека Гумер - Языкознание. Лингвистика. Филология. Языкознание
7	http://superlinguist.ru/	Superlinguist – электронная научная библиотека, посвященная теоретическим и прикладным вопросам лингвистики, а также изучению различным языков
8	http://www.multitran.ru/	Информационная справочная система МультиТран «Электронные словари»
9	Электронно-библиотечная система IPRbooks http://www.iprbookshop.ru/	Электронно-библиотечная система IPRbooks — научно-образовательный ресурс для решения задач обучения в России и за рубежом. Уникальная платформа ЭБС IPRbooks объединяет новейшие информационные технологии и учебную лицензионную литературу. Контент ЭБС IPRbooks отвечает требованиям стандартов высшей школы, СПО, дополнительного и дистанционного образования. ЭБС IPRbooks в полном объеме соответствует требованиям законодательства РФ в сфере образования
10	Издательство «Лань» Электронная библиотечная система http://lanbook.com	Электронно-библиотечная система «Лань» – это ресурс, включающий в себя как электронные версии книг издательства «Лань» и других ведущих издательств учебной литературы, так и электронные версии периодических изданий по естественным, техническим и гуманитарным наукам
11	ЭБС ЮРАЙТ https://www.biblio-online.ru/	Фонд электронной библиотеки составляет более 4000 наименований и постоянно пополняется новинками, в большинстве своем это учебники и учебные пособия для всех уровней профессионального образования от ведущих научных школ с соблюдением требований

10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

При изучении данной дисциплины используются:

1. Компьютерное оборудование лаборатории фонетики
2. Приложение к OpenOffice.org Calc со встроенными формулами.
3. Наборы тестов и контрольных заданий.

Материально-техническая база обеспечивает проведение всех видов деятельности по дисциплине и соответствует действующим санитарным и противопожарным правилам и нормам.

Самостоятельная работа обучающихся в помещениях, оснащенных компьютерной техникой с возможностью подключение к сети Интернет и обеспечением доступа в электронную и информационно-образовательную среду университета.